Early days of message-passing computing: transputers, occam and all that

Tony Hey
Chief Data Scientist
STFC Rutherford Appleton Laboratory
Harwell Campus, UK

The Beginnings

- In 1981 I was on sabbatical at Caltech – as a theoretical particle physicist – and Geoffrey Fox and I went to a colloquium by Carver Mead …
- Carver demonstrated that there were no engineering obstacles to chips getting smaller and faster for the next 20 years
- I went back to the UK and built message-passing machines using the Inmos Transputer
- Geoffrey Fox collaborated with Chuck Seitz in building a hypercube message-passing machine that was usable for scientific applications

The Caltech Cosmic Cube

- Designed and built in the early 1980's by Geoffrey Fox and Chuck Seitz and their teams in Physics and CS
- Processors at nodes of hypercube; message passing between nodes
- Experimented with parallelizing a whole set of scientific applications
- Developed 'Crystalline OS' – CrOS – which was really a library of communication routines
- Demonstrated advantages of virtual addresses, virtual communication channels and kernel-like support at each node

Lessons learnt

- Exploited data parallelism of regular problems by 'domain decomposition'
- For high efficiency, need for lightweight kernels on nodes that allowed for low latency message start up times
- Laid the foundations for parallel programming methodology and parallel performance analysis that are still relevant today
- Irregular problems were more difficult …

The Parallel Computing Landscape (1)

- The Intel Personal Supercomputer
  - iPSC1 had OSF Mach OS on each node which had very high latency for initiating communications
  - iPSC2 released soon afterwards with NX 'Distributed Process' environment based on Caltech's 'Reactive Kernel' OS
- The Transputer Supernode machine
  - Based on the Inmos T800 transputer that combined CPU, FPU, memory and communication channels on chip
  - Native programming language was 'occam', a realization of a simplified version of Hoare's CSP
  - EU 'Supernode' project: Machines manufactured by TelMat and Parsys

The Parallel Computing Landscape (2)

- Many other vendors of parallel message-passing machines:
  - nCUBE
  - Meiko CS-1 and CS-2
  - Suprenum
  - Parsytec
  - IBM SP series
  - …

➢ Each vendor had proprietary message passing system

Portable Message Passing Interfaces?

- The PARMACS macros from the Argonne team
  - Rusty Lusk et al. ‘Mark 1’
- The p4 parallel programming system
  - Rusty Lusk et al. ‘Mark 2’
- The Parallel Virtual Machine PVM
  - Vaidy Sunderam, Al Geist and others
  - Supported message passing across heterogeneous distributed systems
- The PARMACS message passing libraries
  - Developed by Rolf Hempel and others in the EU ‘PPPE’ project

The Origins of MPI (1)

- In 1991 Geoffrey Fox and Ken Kennedy started a community process towards a data parallel Fortran standard
  - This became the High Performance Fortran effort and typified the 'heroic' compiler school of parallel programming
- However, what was clearly needed was a lower level standard for portability of message passing programs across different parallel computers
  - The US were using p4 and Express
  - The EU were using PARMACS in the PPPE and RAPS projects
  - PVM was widely used for programming networks of workstations but not optimized for more closely coupled parallel machines

The Origins of MPI (2)

- Workshop on Standards for Message Passing in a Distributed Memory Environment
  - Williamsburg, Virginia, April 1992
  - Organized by Jack Dongarra and David Walker
  - Sponsored by CRPC and Ken Kennedy urged action
- In summer of 1992, I contacted Jack Dongarra about starting such a standardization activity
  - Did not want US and Europe to diverge
  - Co-wrote a first draft of an MPI standard with Jack Dongarra, Rolf Hempel and David Walker in October 1992, now known as MPI-0

The Origins of MPI (3)

Organized BOF session at SuperComputing 92 in Minneapolis

- MPI-0 document served as a catalyst
- Marc Snir of IBM emailed me to say 'he was happy to have been plagiarized'
- I have no idea why we left the obvious collective communications routines out of MPI-0
- Rusty Lusk and Bill Gropp from Argonne volunteered to produce an open source implementation of the evolving MPI standard
- And the EU PPPE project paid for the beer …

The MPI Process

- Followed procedures of HPF Forum
  - Set ambitious goal of agreeing a standard within one year
- Met every 6 weeks in Dallas airport hotel
  - I sent Ian Glendinning from my group in Southampton funded by EU PPPE project
- In my opinion MPI-1 succeeded because:
  - Argonne produced an open source implementation
  - Exceptional technical leadership from people like Marc Snir from IBM and Jim Cownie from Meiko
  - It was needed and had the support of the community

Parkbench: Portable DM Message-Passing Kernels and Benchmarks

- Advent of MPI meant that it was possible to assemble suite of Message-Passing benchmarks for performance analysis of machines and applications
- EU Genesis project defined 3 levels of benchmarks
  - Low-level, Kernels and a set of Compact Applications implemented with PARMACS libraries
- International Parkbench Group
  - Combined Genesis methodology with Linear Algebra and NAS Parallel Benchmarks implemented with MPI-1

➢ But the marketing community preferred Jack's Top500 Benchmark …

Acknowledgements

- Many thanks to Jack Dongarra, Rolf Hempel and David Walker

- A useful ‘aide memoire’ was the article by Dongarra, Fagg, Hempel and Walker in the Encyclopedia of Electronics and Electrical Engineering (Wiley)

MPI-0 Reference

Jack Dongarra, Rolf Hempel, Tony Hey, David Walker

'A Draft Standard for Message Passing on Distributed Memory Computers'

Proceedings of the Fifth ECMWF Workshop on the Use of Parallel Processors in Meteorology:

"Parallel Supercomputing in Atmospheric Science"
Editors: Geerd-R Hoffmann and Tuomo Kauranne
Published by World Scientific, 1993